

(12)特許協力条約に基づいて公開された国際出願

(19) 世界知的所有権機関
国際事務局(43) 国際公開日
2003年11月27日 (27.11.2003)

PCT

(10) 国際公開番号
WO 03/098597 A1

- (51) 国際特許分類: G10L 11/00, 13/06
- (21) 国際出願番号: PCT/JP03/01954
- (22) 国際出願日: 2003年2月21日 (21.02.2003)
- (25) 国際出願の言語: 日本語
- (26) 国際公開の言語: 日本語
- (30) 優先権データ:
特願2002-141390 2002年5月16日 (16.05.2002) JP
- (71) 出願人 (米国を除く全ての指定国について): 科学技術振興事業団 (JAPAN SCIENCE AND TECHNOLOGY CORPORATION) [JP/JP]; 〒332-0012 埼玉県川口市本町4丁目1番8号 Saitama (JP). 株式会社国際

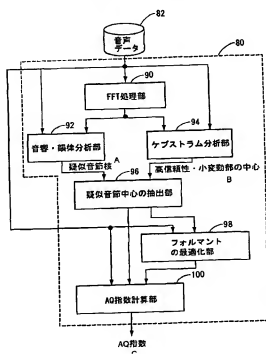
電気通信基礎技術研究所 (ADVANCED TELECOMMUNICATION RESEARCH INSTITUTE INTERNATIONAL) [JP/JP]; 〒619-0288 京都府相楽郡精華町光台二丁目2番地2 Kyoto (JP).

- (72) 発明者: および
(75) 発明者/出願人 (米国についてのみ): キャンベル ニック (CAMPBELL, Nick) [GB/JP]; 〒619-0288 京都府相楽郡精華町光台二丁目2番地2 C/O ATR Kyoto (JP).
モクタリ パーハム (MOKHTARI, Parham) [AU/JP]; 〒619-0288 京都府相楽郡精華町光台二丁目2番地2号 C/O ATR Kyoto (JP).
- (74) 代理人: 清水 敏 (SHIMIZU, Satoshi); 〒530-0047 大阪府大阪市北区西天満2丁目3番9号 オーク西天満ビル3階 Osaka (JP).
- (81) 指定国 (国内): CA, US.

[続表有]

(54) Title: SYLLABIC KERNEL EXTRACTION APPARATUS AND PROGRAM PRODUCT THEREOF

(54) 発明の名称: 音節核の抽出装置およびそのプログラム製品



- 92... SPEECH DATA
90... FFT PROCESSOR
92... ACOUSTIC/RHYTHM ANALYZER
94... CEPSTRUM ANALYZER
96... PSEUDO SYLLABIC KERNEL
96... CENTER OF HIGH RELIABILITY/SMALL FLUCTUATION
98... PSEUDO SYLLABIC CENTER EXTRACTOR
98... FORMANT OPTIMIZER
100... AQ INDEX CALCULATOR
C... AQ INDEX

(57) Abstract: An apparatus which automatically determines a portion showing characteristic of a speech waveform with a high reliability includes an acoustic/rhythm analyzer (92) for calculating distribution of energy of a predetermined frequency region of a speech waveform from data on the time axis and extracting a region where the syllable of the speech waveform is pronounced stably according to the distribution and the speech waveform pattern, a cepstrum analyzer (94) for estimating a region where the speech waveform change is preferably controlled by a speaker according to the distribution of the spectrum of the speech waveform on the time axis, and a pseudo syllabic center extractor (96) for deciding the region extracted as a region pronounced stably and estimated that the change is preferably controlled by the speaker as a highly reliable portion of the speech waveform.

(57) 要約: 音声波形の特徴を高い信頼性で示す部分を自動的に決定できるようにするための装置は、データから音声波形のうちの所定周波数領域のエネルギーの時間軸上の分布を算出し、当該分布および音声波形のピッチに基づいて、音声波形の各節のうち安定して発声されている領域として抽出する音響・韻律分析部 (92) と、音声波形のスペクトルの時間軸上の分布に基づいて、音声波形のうちの、その変化が話者により良好に制御されている領域を推定するケプストラム分析部 (94) と、安定して発声されている領域として抽出され、かつ話者によってその変化が良好に制御されていると推定された領域を音声波形の高信頼性部分として決定する疑似音節核の抽出部 (96) とを含む。



規則4.17に規定する申立て:

- CAの指定のための不利にならない開示又は新規性喪失の例外に関する申立て (規則4.17(v))
- CAの指定のための不利にならない開示又は新規性喪失の例外に関する申立て (規則4.17(v))
- USの指定のための不利にならない開示又は新規性喪失の例外に関する申立て (規則4.17(v))
- USの指定のための不利にならない開示又は新規性喪失の例外に関する申立て (規則4.17(v))

添付公開書類:

- 国際調査報告書
- 不利にならない開示又は新規性喪失の例外に関する申立て

2文字コード及び他の略語については、定期発行される各PCTガゼットの巻頭に掲載されている「コードと略語のガイダンスノート」を参照。

明細書

音節核の抽出装置およびそのプログラム製品

技術分野

この発明は、一般的には音声波形からその特徴を高い信頼性で示す部分を抽出
5 するための技術に関し、特に、音声波形の発生源の状態を高い信頼性で推定する
ために有効な領域を、音声波形から抽出するための技術に関する。

背景技術

[用語の定義1]

最初に、この節で使用される用語について定義する。

10 「緊張音」(pressed sound)とは、発声の際に声門が緊張して
いるために声門を気流が通過しにくく、かつ通過をする際の気流の加速度が大き
くなるように発声される音のことをいう。この場合、声門気流波形はサインカー
ブから大きく変形し、その微分波形の傾きが局部的に大きくなる。音声がかうし
た特徴を有する場合、「緊張性」の音声であると呼ぶことにする。

15 「氣息音」(breathy sound)とは、発声の際に声門に緊張がな
いたために気流が通過しやすく、その結果声門気流波形がサインカーブに近くなる
ように発声される音をいう。この場合、声門気流波形の微分波形の傾きが局部的
に大きくなることはない。音声がかうした特徴を有する場合、「氣息性」の音声
であると呼ぶことにする。

20 「地声」(モーダル、modal)とは、緊張音と氣息音との中間の発声のこ
とをいう。

「AQ指数」(Amplitude Quotient)とは、声門(声帯)
気流の波形のピークツーピークの振幅を、声門気流の波形の微分の振幅の最小値
で除した値のことをいう。

25 [従来技術]

音声認識と並んで重要な音声研究分野に、音声合成がある。最近の信号処理技術の発達により、音声合成が既に多くの分野で利用されている。しかし、今までの音声合成は単にテキスト情報を音声化しているだけとはいえ、人間が発話する際のような微妙な感情の表現までは行なえない。

たとえば、人間が発話する際には、怒り、喜び、および悲しみなどの情報が、発話内容以外の情報、つまり声色などにより伝達される。このように発話に付随する、言語以外の情報をパラ言語情報と呼ぶ。これらはテキスト情報のみでは表わせない情報である。しかし従来の音声合成では、こうしたパラ言語情報を伝達することは難しかった。マンマシンインタフェースをより効率的なものとするためには、テキスト情報だけではなくパラ言語情報も音声合成の際に伝えられるようにすることが望ましい。

こうした問題を解決するために、種々の発話スタイルで連続的に音声合成を行なおうとする試みがある。ひとつの具体的な方策として次のようなものがある。すなわち、発話を録音してデータ処理可能な形でデータベース化し、さらにその中で所望の特徴(怒り、喜び、悲しみなど)を表わすと思われる発話単位にそれらの特徴を示すラベルを付ける。音声合成の際には所望のパラ言語情報に対応したラベルが付けられた音声を利用する。

しかし、十分な広さの発話スタイルをカバーできるようにデータベースを構築しようとするれば、膨大な量の録音音声进行处理しなければならない。そのために、自動的にオペレータの介入なく確実にそうした特徴の抽出とラベル付け処理とを行なえるようにする必要がある。

以下、パラ言語情報の一例を挙げる。発話スタイルの一つとして、緊張音と氣息音という区別がある。緊張音では声門が緊張しているために、どちらかという強い発声となる。一方氣息音では、音声はサインカーブに近く、強いという印象はない。したがって緊張音と氣息音という区別も発話スタイルの一つとして重要であり、その程度を数量化できれば、パラ言語情報として利用できる可能性がある。

緊張音と氣息音との音質を区別する音響学的な指標については、今までにも数多くの研究がなされてきた。たとえば文末にリストした参考文献1を参照された

い。しかし、そうした研究の多くは、持続的に安定して母音を発音している間に録音された発話（または歌）を対象としたものに限定されていた。実際、膨大な量の発話の録音データから得られた音響測定データに基づいて、緊張性と気息性との程度を信頼性高く計量しなければならないというのは非常に大きな問題であり、かつ実現された場合には非常に有用となるであろう。

スペクトルドメインでの音源の属性を推定しようとする様々な手段が提案されて来たが、それよりも直接的な推定が、声門気流の波形とその導関数との組み合わせによって得られるはずである。そうした推定の一例が文末の参考文献2において提案されたAQ指数である。

参考文献2では、AQ指数の一つの利点として、音圧レベル（SPL）から比較的独立していること、およびその値が主として発音の質的なものに依存していることがあげられている。他の利点として考えられるのは、このパラメータが純粋に振幅ドメインのものであって、種々の発話スタイルに応じた、推定された声門気流の時間ドメインの特徴量を測定する際の誤差源に対して比較的免疫性があることである。また、参考文献2の著者らによれば、様々な発音スタイルで「a」という母音を持続して発音した場合、4人の男性と4人の女性との全てに対して、発音を気息性のものから緊張性に変えていくにしたがって、AQ指数の値は単調に減少したとのことである（参考文献2の第136頁）。したがってAQ指数は、ここで我々が解決しようとしている問題に関して有効である可能性が高い。ただし、AQ指数が有効となるためには、次の条件が満足される必要がある。

1) 録音された通常の発話について、ロバストでかつ信頼性高くAQ指数を測定できること、および

2) そうした条件で測定された知覚上の特徴が顕著な部分を確認することができること。

このような条件を満足させるためには、自然に発話された音声などの物理量を表わす音声波形から、いかにして信頼性高く音声波形の特徴を表わすパラメータを抽出できるかが重要である。特に音声の場合のように、発話が話者によりその細部まで完全にはコントロールされているわけでない場合、また様々な人が様々

なスタイルで発話する場合には、パラメータを抽出すべき部分として信頼性がおける場所と、そうでない場所とが存在することが考えられる。そのため、音声波形のうちのどの部分を処理対象とするかが重要である。またそのために、日本語のように音節が発音の単位となる場合、音節の中心部（仮にこれを「音節核」と呼ぶ。）を誤りなく抽出できるようにすることが必要である。

発明の開示

したがって、本発明の目的は、音声波形の特徴を高い信頼性で示す部分を自動的に処理で決定することを可能とすることである。本発明の他の目的は、音声波形の特徴を高い信頼性で示す部分をオペレータの介在なく決定することを可能とすることである。本発明のさらに他の目的は、音節核を高い信頼性で自動的に抽出できるようにすることである。

本発明の第1の局面は、複数の節に分解可能な、物理的量を表わす音声波形のデータに基づいて、音声波形の特徴を高い信頼性で示す部分を決定するための装置と、そうした装置としてコンピュータを動作させるプログラムに関する。この装置は、データから音声波形のうちの所定周波数領域のエネルギーの時間軸上の分布を算出し、当該分布および音声波形のピッチに基づいて、音声波形の各節のうち、音声波形の発生源によって安定して発生されている領域を抽出するための抽出手段と、データから音声波形のスペクトルの時間軸上の分布を算出し、当該スペクトルの時間軸上の分布に基づいて、音声波形のうち、その変化が発生源により良好に制御されている領域を推定するための推定手段と、推定手段の出力と、発生源によって安定して発生されている領域として抽出手段により抽出され、かつ発生源によってその変化が良好に制御されていると推定手段によって推定された領域を音声波形の高信頼性部分として決定するための手段とを含む。

抽出手段による抽出結果と、推定手段による推定結果との双方に基づいて音声波形の高信頼性部分を決定するので、決定結果がより確実なものとなる。

抽出手段は、データに基づいて、音声波形の各区間が有声区間か否かを判定するための有声判定手段と、音声波形の所定周波数領域のエネルギーの時間軸上の分布の波形の極小部で音声波形を節に分離するための手段と、音声波形のうち、

各節内で、当該節内のエネルギーのピークを含み、かつ有声判定手段により有声区間であると判定された区間であって、かつ所定周波数領域のエネルギーが所定のしきい値以上である領域を抽出するための手段とを含んでもよい。

5 有声と判定された区間であって、かつ所定周波数領域のエネルギーが所定のしきい値以上である領域が抽出されるので、発話者が安定して発声している区間を確実に抽出できる。

また好ましくは、推定手段は、音声波形に対する線形予測分析を行ないフォルマント周波数の推定値を出力するための線形予測手段と、データを用いて、線形予測手段によるフォルマント周波数の推定値の非信頼性の時間軸上の分布を算出する第1の算出手段と、線形予測手段の出力に基づいて、音声波形の時間軸上のスペクトル変化の局所的な分散の、時間軸上の分布を算出するための第2の算出手段と、第1の算出手段により算出されたフォルマント周波数の推定値の非信頼性の時間軸上の分布と、第2の算出手段により算出された音声波形のスペクトル変化の局所的な分散の時間軸上の分布との双方に基づいて、音声波形の15 変化が発生源により良好に制御されている領域を推定するための手段とを含む。

フォルマント周波数の推定値の非信頼性と、音声波形の時間軸上のスペクトル変化の局所的な分散との双方に基づいて、音声波形の変化が発生源により良好に制御されている領域が推定される。振動変化の発生源（たとえば発話者）が明確な意図をもって振動を制御している領域が推定できるので、そうした領域から振20 動の特徴量を算出すれば、算出された特徴量の信頼性が高くなることが期待できる。

決定するための手段は、推定手段により音声波形の変化が発生源により良好に制御されていると推定された領域のうち、抽出手段により抽出された領域に含まれる領域を音声波形の高信頼性部分として決定するための手段を含んでもよい。

25 音声波形の変化が発生源により良好に制御されていると推定された領域であって、かつ発生源により音声波形が安定に発生されているもののみを高信頼性部分として決定する。したがって真に信頼性が高い部分を抽出できる。

本発明の他の局面は、音声信号を擬似音節に分離し、さらに各擬似音節の核部分を抽出するための擬似音節核抽出装置と、そうした装置としてコンピュータを

動作させるプログラムとに関する。この擬似音節核抽出装置は、音声信号の各区
間が有声区間か否かを判定するための有声判定手段と、音声信号の所定周波数領
域のエネルギーの時間的な分布の波形の極小部で音声信号を擬似音節に分離する
ための手段と、音声信号のうち、各擬似音節内でのエネルギーのピークを含み、
5 かつ有声判定手段により有声区間であると判定された区間であって、かつ所定周
波数領域のエネルギーが所定のしきい値以上である領域を当該擬似音節の核とし
て抽出するための手段とを含む。

有声区間であると判定された区間であって、かつ所定周波数領域のエネルギー
が所定のしきい値以上である領域が擬似音節の核として抽出されるので、発話者
10 が安定して発声しているときの音声抽出することができる。

本発明のさらに他の局面は、音声信号の特徴を高い信頼性で示す部分を決定す
るための装置と、そうした装置としてコンピュータを動作させるプログラムとに
関する。当該装置は、音声信号に対する線形予測分析を行なうための線形予測手
段と、線形予測手段によるフォルマントの推定値と、音声信号とに基づいて、フ
15 オルマントの推定値の非信頼性の時間軸上の分布を算出するための第1の算出手
段と、線形予測手段による線形予測分析の結果に基づいて、音声信号のスペクト
ル変化の局所的な分散の時間軸上の分布を算出するための第2の算出手段と、第
1の算出手段により算出されたフォルマント周波数の推定値の非信頼性の時間軸
上の分布と、第2の算出手段により算出された音声波形のスペクトル変化の局所
20 的な分散の時間軸上の分布との双方に基づいて、音声波形の変化が発生源により
良好に制御されている領域を推定するための手段とを含む。

フォルマントの推定値の非信頼性の時間軸上の分布も、音声信号のスペクトル
変化の局所的な分散の時間軸上の分布も、その極小部ではいずれも音声信号のう
ちでその発生源により音声波形の発生が良好に制御されている部分を示す。これ
25 らの双方を用いて領域を推定するので、音声波形の発生が良好に制御されている
部分を信頼性高く特定することができる。

図面の簡単な説明

第1図は、本発明の一実施例のプログラムを実行するコンピュータシステムの

外観を示す図である。

第2図は、第1図に示すコンピュータシステムのブロック図である。

第3図は、本発明の一実施例のプログラムの全体構成をブロック図形式で示す図である。

5 第4図は、音声データの構成を模式的に示す図である。

第5図は、第3図に示す音響・韻律分析部92のブロック図である。

第6図は、第3図に示すケプストラム分析部94のブロック図である。

第7図は、第6図に示す規格化および統合部144のブロック図である。

第8図は、第3図に示すフォルマントの最適化部98のブロック図である。

10 第9図は、第3図に示すAQ指数計算部100のブロック図である。

第10図は、本発明の一実施例のプログラムによる表示例を示す図である。

第11図は、音声データのうち、緊張音と判断される一時点での声門気流波形の推定値、声門気流波形の微分の推定値、および推定された声門気流波形のスペクトルを示す図である。

15 第12図は、音声データのうち、氣息音と判断される一時点での声門気流波形の推定値、声門気流波形の微分の推定値、および推定された声門気流波形のスペクトルを示す図である。

第13図は、感知された氣息性と音響的に測定されたAQ指数との間の関連を示す散布図である。

20 発明を実施するための最良の形態

以下に述べる本発明の実施例は、コンピュータおよびコンピュータ上で動作するソフトウェアにより実現される。もちろん、以下に述べる機能の一部又は全部を、ソフトウェアでなくハードウェアで実現することも可能である。

[用語の定義2]

25 以下、本実施例の説明で使用される用語について定義する。

「擬似音節」とは、音声信号から所定の信号処理によって決定される信号の切れ目のことを指し、日本語音声の場合の音節を推定したものに対応する。

「ソノラントエネルギー」とは、音声信号のうちで、所定周波数（たとえば6

0 Hz ~ 3 kHz の周波数領域) のエネルギーのことをいい、デシベルで表わされる。

「信頼性の中心」 (center of reliability) とは、音声波形に対する信号処理の結果、音声波形のうちで、対象となる音声波形の特徴を信頼性高く抽出することができることとみなされることとなった領域のことをいう。

「ディップ」とは、グラフなどの図形がくびれた部分のことをいう。特に、時間の関数として変化するような値の時間軸上の分布により形成される波形のうち、極小値に対応する部分をいう。

「非信頼性」とは、信頼性のなさを表わす尺度のことをいう。非信頼性は信頼性の逆の概念である。

第1図に、本実施例で利用されるコンピュータシステム20の外観図を、第2図にコンピュータシステム20のブロック図を、それぞれ示す。なおここに示すコンピュータシステム20はあくまで一例であり、この他にも種々の構成が可能である。

第1図を参照して、コンピュータシステム20は、コンピュータ40と、いずれもこのコンピュータ40に接続されたモニタ42、キーボード46、およびマウス48を含む。コンピュータ40にはさらに、CD-ROM (Compact Disc Read-Only Memory) ドライブ50と、FD (Flexible Disk) ドライブ52とが内蔵されている。

第2図を参照して、コンピュータシステム20はさらに、コンピュータ40に接続されるプリンタ44を含むが、これは第1図には示していない。またコンピュータ40はさらに、CD-ROMドライブ50およびFDドライブ52に接続されたバス66と、いずれもバス66に接続された中央演算装置 (Central Processing Unit: CPU) 56、コンピュータ40のブートアッププログラムなどを記憶したROM (Read-Only Memory) 58、CPU56が使用する作業エリアおよびCPU56により実行されるプログラムの格納エリアを提供するRAM (Random Access Memory) 60、および後述する音声データベースを格納したハードディスク54を含む。

以下に述べる実施例のシステムを実現するソフトウェアは、たとえば、CD-ROM 6 2 のような記録媒体上に記録されて流通し、CD-ROM ドライブ 5 0 のような読取装置を介してコンピュータ 4 0 に読み込まれ、ハードディスク 5 4 に格納される。CPU 5 6 がこのプログラムを実行する際には、ハードディスク 5 4 からこのプログラムを読み出して RAM 6 0 に格納し、図示しないプログラムカウンタによって指定されるアドレスから命令を読み出して実行する。CPU 5 6 は、処理対象のデータをハードディスク 5 4 から読み出し、処理結果を同じくハードディスク 5 4 に格納する。

コンピュータシステム 2 0 の動作自体は周知であるので、ここではその詳細に

ついては繰り返さない。

なお、ソフトウェアの流通形態は上記したように記憶媒体に固定された形には限定されない。たとえば、ネットワークを通じて接続された他のコンピュータからデータを受取る形で流通することもあり得る。また、ソフトウェアの一部が予めハードディスク 5 4 中に格納されており、ソフトウェアの残りの部分をネットワーク経由でハードディスク 5 4 に取込んで実行時に統合するような形の流通形態もあり得る。

一般的に、現代のプログラムはコンピュータのオペレーティングシステム (OS) によって提供される汎用の機能を利用し、それらを所望の目的にしたがって組織化した形態で実行することにより前記した所望の目的を達成する。したがって、以下に述べる本実施例の各機能のうち、OS またはサードパーティが提供する汎用的な機能を含まず、それら汎用的な機能の実行順序の組合せだけを指定するプログラム (群) であっても、それらを利用して全体的として所望の目的を達成する制御構造を有するプログラム (群) である限り、それらが本発明の技術的範囲に含まれることは明らかである。

本実施例のプログラムを装置とみなして機能的に示したのが第 3 図以下のブロック図である。第 3 図を参照して、この装置 8 0 は、ハードディスク 5 4 に格納された音声データ 8 2 に対して以下に説明する処理を行なって、音声データに含まれる各処理単位 (たとえば音節) ごとに前述した A Q 指数を算出し出力するためのものである。なお、音声データは後述するように 1 フレーム 3 2 m s e c と

なるように予めフレーム化されている。

装置 80 は、音声データに対して高速フーリエ変換 (Fast Fourier Transform: FFT) を行なう FFT 処理部 90 と、FFT 処理部 90 の出力を用い、音声データにより表わされる音声波形のうちの 60 Hz ~ 3 kHz の周波数領域のエネルギーの時間的変化および音声のピッチの変化に基づいて、音声データにより表わされる音声波形の各音節のうち、話者の発声機構によって安定して発生されている領域 (これを以後「擬似音節核」と呼ぶ。) とを抽出する音響・韻律分析部 92 と、音声データ 82 に対してケプストラム分析を行ない、さらに、FFT 処理部 90 の出力を用いてケプストラム分析の結果音声スペクトルの変化が少なく、音声データの特徴を信頼性高く抽出できると思われる部分 (これを「高信頼性・小変動部の中心」または「高信頼・小変動の中心」または単に「信頼性の中心」と呼ぶ。) を推定するためのケプストラム分析部 94 とを含む。

装置 80 はさらに、ケプストラム分析部 94 の出力する高信頼性・小変動部の中心の中で、音響・韻律分析部 92 の出力する擬似音節核の中にあるものだけを擬似音節中心として抽出するための擬似音節中心の抽出部 96 と、擬似音節中心の抽出部 96 によって抽出された擬似音節中心に対応する音声データに対して、フォルマントの初期推定と最適化処理とを行なって最終的なフォルマントの推定値を出力するためのフォルマントの最適化部 98 と、音声データに対して、フォルマントの最適化部 98 から出力されるフォルマント値を用いた適応的フィルタ処理などの信号処理を行なって声門気流波形の微分を推定し、さらにそれを積分することによって声門気流波形を推定し、それらに基づいて AQ 指数を計算するための AQ 指数計算部 100 とを含む。

第 4 図は、音声データの構成を模式的に示す図である。第 4 図を参照して、音声データ波形 102 は、それぞれ 32 msec ごとのフレームに分けられ、かつ前後のフレーム間では 8 msec ごとにずらしてデジタル化されている。そして、後述する処理では、たとえばある時点 t_0 では第 1 のフレームを先頭として処理をし、次の時点 t_1 では 8 msec ずれた次の第 2 のフレームを先頭として処理をする、という形で処理を行っていく。

第5図は、第3図に示す音響・韻律分析部92のブロック図である。第5図を参照して、音響・韻律分析部92は、音声波形から測定される音源のピッチを用いて、処理対象のフレームが有声区間か否かを判定する（この方法については参考文献3を参照）ためのピッチ判定部110と、FFT処理部90の出力に基づいて所定周波数領域（60Hz～3kHz）のソノラントエネルギーの時間軸上の波形分布を算出するためのソノラントエネルギー算出部112と、ソノラントエネルギー算出部112によって算出されるソノラントエネルギーの時間軸上の分布波形の輪郭に対して凸包アルゴリズムを適用することにより、ソノラントエネルギーの時間軸上の分布波形の輪郭の中のディップを検出して、入力音声に擬似音節に分割する（この方法については参考文献4および5を参照）ためのディップ検出部114と、ディップ検出部114によって得られた擬似音節中の、ソノラントエネルギーの最大値（SE_{peak}）が得られる点を起点として、その左右に、ソノラントエネルギーが所定のしきい値（ $0.8 \times SE_{peak}$ ）より大きく、かつピッチ判定部110によって有声区間であると判定されたフレームであって、かつ同じ擬似音節中のフレームを1フレームずつ広げていくことにより、擬似音節核を出力するための有声・エネルギー判定部116とを含む。

第6図は、第3図に示すケプストラム分析部94のブロック図である。第6図を参照して、ケプストラム分析部94は、音声データ82の音声波形に対して選択的線形予測（Selective Linear Prediction: SLP）分析を行なって、SLPケプストラム係数 $c_{r,i}$ を出力するための線形予測分析部130と、このケプストラム係数に基づいて先頭の4つのフォルマントの周波数と帯域との初期推定値を算出するためのフォルマント推定部132とを含む。フォルマント推定部132は、参考文献6により提案された線形ケプストラム-フォルマントマッピングを利用し、かつ同じデータのサブセットを使用して注意深く測定された母音フォルマントに対するマッピングを学習させてある。この学習については、参考文献7を参照されたい。

ケプストラム分析部94はさらに、推定されたフォルマント周波数などに基づいてケプストラム係数 C_i^{sine} を再計算するためのケプストラム再生部136と、FFT処理部90の出力に対して対数変換およびコサイン逆変換（IDCT）を

行なってFFT係数を算出するための対数変換および逆DCT部140と、ケプストラム再生部136により計算されたケプストラム係数 C_i^{imp} と、対数変換および逆DCT部140により計算されたFFT係数 C_i^{FFT} との間の差を表わす値として次の式により定義されるケプストラム距離 d_i^2 を計算し、フォルマント推定部132によって推定されたフォルマント周波数などの値の非信頼性を表わす指標として出力するためのケプストラム距離計算部142とをさらに含む。

$$d_i^2 = \text{Sum}_i \{ i^2 \cdot (c_i^{imp} - c_i^{FFT})^2 \} \quad (1)$$

フォルマント推定部132、ケプストラム再生部136、ケプストラム距離計算部142、および対数変換および逆DCT部140により、線形予測分析の結果に基づいて推定されたフォルマント周波数などの値の非信頼性が算出される。

ケプストラム分析部94はさらに、線形予測分析部130の出力するケプストラム係数から Δ ケプストラムを算出する為の Δ ケプストラム算出部134と、 Δ ケプストラム算出部134の出力する Δ ケプストラムに基づいて、各フレームごとに、そのフレームを含む5フレームのスペクトル変化の大きさの分散を算出する為のフレーム間分散算出部138とを含む。フレーム間分散算出部138の出力は、局所的なスペクトルの動きの時間軸上の分布波形の輪郭を表わすものとなり、その極小値は、参考文献8で提案されている調音音声学理論にならって例えば、制御された動きCM (Controlled Movement) を示すものと考えることができる。

さらにケプストラム分析部94は、ケプストラム距離計算部142の出力するフォルマント周波数の推定値の非信頼性を示す値と、フレーム間分散算出部138の出力する各フレームごとの局所的なフレーム間分散値を受け、両者の値を規格化し統合して、フレームごとの音声信号の非信頼性を示す値の時間軸上の分布波形として出力するための規格化および統合部144と、規格化および統合部144の出力する非信頼性の値の時間軸上の分布波形により形成される波形の輪郭のディップを凸包アルゴリズムにより検出して、信頼性の中心候補として出力するための信頼性の中心候補出力部146とを含む。

第7図は、第6図に示す規格化および統合部144のブロック図である。第7

図を参照して、規格化および統合部 144 は、ケプストラム距離計算部 142 に
より出力されたケプストラム距離を [0, 1] の値に規格化するための第 1 の規
格化部 160 と、フレーム間分散算出部 138 が各フレームごとに算出するフレ
ーム間分散の値を [0, 1] の値に規格化するための第 2 の規格化部 162 と、
局所的なフレーム間分散の値の時間軸上の位置を、ケプストラム距離計算部 14
2 の出力するケプストラム距離のサンプリングタイミングと一致させるように線
形補間処理を行なうための補間処理部 164 と、第 1 の規格化部 160 の出力と
補間処理部 164 の出力とを 1 フレームごとに平均して出力するための平均計算
部 166 とを含む。平均計算部 166 の出力は、統合された値の時間軸上の分布
波形の輪郭を表わす。信頼性の中心候補出力部 146 によってこの波形の輪郭の
ディップ（極小部）を検出することにより、非信頼性が最も低い部分（信頼性が
最も高い部分）を信頼性の中心の候補として特定することができる。

第 8 図は、第 3 図に示すフォルマントの最適化部 98 のブロック図である。第
8 図を参照して、フォルマントの最適化部 98 は、音声波形に対して FFT 処理
を行なうための FFT 処理部 180 と、FFT 処理部 180 の出力に対して対数
変換およびコサイン逆変換を行なうための対数変換および逆 DCT 部 182 と、
対数変換および逆 DCT 部 182 の出力する FFT ケプストラム係数と、後述す
るフォルマントの推定値との間の距離を計算するためのケプストラム距離計算部
184 と、信頼性の中心候補の各々における第 1 ～第 4 のフォルマント周波数の
初期推定値を初期値とし、ケプストラム距離計算部 184 が計算する距離を最小
にするように山登り法によってフォルマントの推定値を最適化するための距離最
小化処理部 186 とを含む。距離最小化処理部 186 によって最適化されたフォ
ルマント推定値がフォルマントの最適化部 98 の出力として AQ 指数計算部 10
0 に与えられる。

第 9 図を参照して、AQ 指数計算部 100 は、音声信号のうちで音節中心に相
当する位置の 64 msec の部分のうち、70 Hz 以上の周波数成分のみを選択
的に通過させるためのハイパスフィルタ 200 と、ハイパスフィルタ 200 の出
力のうち、最適化された第 4 フォルマント周波数とその帯域との和以下の周波数
成分のみを選択的に通過させるための適応的ローパスフィルタ 202 と、適応的

ローパスフィルタ 202 の出力に対し、第 1 ～ 第 4 フォルマント周波数を用いた適応的逆フィルタ処理を行なうための適応的逆フィルタ 204 とを含む。適応的逆フィルタ 204 の出力は、声門気流波形の微分波形となる。

AQ 指数計算部 100 はさらに、適応的逆フィルタ 204 の出力を積分して声門気流波形を出力するための積分回路 206 と、積分回路 206 の出力のピークツーピークの最大振幅を検出するための最大ピーク間振幅検出回路 208 と、適応的逆フィルタ 204 の出力の負のピークの最大振幅を検出するための最大の負のピーク振幅検出回路 210 と、最大の負のピーク振幅検出回路 210 の出力に対する最大ピーク間振幅検出回路 208 の出力の比を算出するための比計算回路 212 とを含む。比計算回路 212 の出力が AQ 指数である。

以上に説明した装置は以下のように動作する。まず、使用された音声データ 82 について説明する。この音声データは参考文献 9 で使用されたものであり、日本語のネイティブスピーカである女性の 3 つの物語の朗読を録音して作成されたものである。この物語は、怒りと、喜びと、悲しみという感情を引き起こすように予め作成されていたものである。物語の各々は 400 文の長さ（おおよそ 30,000 音素）以上の発話を含む。各発話は別々の音声波形ファイルに格納され処理された。

各文の発話データは FFT 処理部 90 による FFT 処理の後、以下のようにして処理される。処理は大きく見て二つの系統に分かれ実行される。第 1 の系統は音響・韻律分析部 92 で行なわれる音響韻律的な処理であり、他の系統はケプストラム分析部 94 が行なう音響音声学的な処理である。

音響韻律的な系統の処理では、第 5 図に示すソノラントエネルギー算出部 112 によって 60 Hz ～ 3 kHz 周波数領域のソノラントエネルギーが算出される。ソノラントエネルギー算出部 112 の出力する一文の発話データの全体波形の輪郭から、ディップ検出部 114 が凸包アルゴリズムによりディップを検出する。このディップにより、この発話文は擬似音節に分割される。

有声・エネルギー判定部 116 は、擬似音節の中でソノラントエネルギーが最大（S E peak）となる点を見つける。この点が擬似音節核の初期点である。有声・エネルギー判定部 116 はさらに、この擬似音節核の初期点から始めて、その左右

5 に向かい、ソノラントエネルギーが $0.8 \times S E_{peak}$ 以下のフレーム、またはピッチ判定部 110 が有声でないと判定したフレーム、または擬似音節の外のフレームに出会うまで、擬似音節核の範囲を広げる。こうして擬似音節核の境界が決定される。この情報は擬似音節中心の抽出部 96 に与えられる。なお、ここでしきい値として 0.8 の値を用いているが、これは単なる例であって、応用によりこのしきい値を適切な値に代える必要がある。

10 第6図を参照して、入力された一つの発話文に対して線形予測分析部 130 が線形予測分析を行ない、SLP ケプストラム係数を出力する。 Δ ケプストラム算出部 134 がこの SLP ケプストラム係数に基づいて Δ ケプストラムを算出し、フレーム間分散算出部 138 に与える。フレーム間分散算出部 138 は、この Δ ケプストラム係数に基づき、各フレームごとに、そのフレームを含む 5 フレームの中での局所的なスペクトル変化の分散を計算する。この分散が小さいほど発話者の発声が発話者によりよく制御されていると考えられ、逆にこの分散が大きいと話者による制御がよくされていないと考えられるので、フレーム間分散算出部 138 の出力は発話者の発声が信頼されない程度（非信頼性）を表わすと考えられる。

20 第6図をさらに参照して、フォルマント推定部 132 は、線形ケプストラムフォルマントマッピングを用い、SLP ケプストラム係数に基づいて第1～第4フォルマントの周波数と帯域とを推定する。ケプストラム再生成部 136 は、フォルマント推定部 132 により推定された第1～第4フォルマントに基づいて逆にケプストラム係数を算出しケプストラム距離計算部 142 に与える。対数変換および逆 DCT 部 140 は、フォルマント推定部 132 およびケプストラム再生成部 136 が処理したのと同じフレームのものと音声データに対して対数変換およびコサイン逆変換を行なって FFT ケプストラム係数を算出しケプストラム距離計算部 142 に与える。ケプストラム距離計算部 142 は、ケプストラム再生成部 136 からのケプストラム係数と対数変換および逆 DCT 部 140 からのケプストラム係数との間の距離を前述の式 (1) にしたがって計算する。この結果得られるのは、フォルマント推定部 132 が推定したフォルマントの非信頼性を示す値の時間軸上の分布を表わす波形と考えられる。ケプストラム距離計算部 14

2 は、この結果を規格化および統合部 1 4 4 に与える。

第 7 図を参照して、規格化および統合部 1 4 4 の第 1 の規格化部 1 6 0 は、第 6 図のケプストラム距離計算部 1 4 2 の出力する、フォルマントの推定値から算出された各フレームごとの非信頼性値を [0, 1] の範囲に正規化して平均計算部 1 6 6 に与える。第 2 の規格化部 1 6 2 は、第 6 図のフレーム間分散算出部 1 3 8 が出力する、フレームごとに計算された局所的なフレーム間分散の値を [0, 1] の範囲に正規化して補間処理部 1 6 4 に与える。補間処理部 1 6 4 は、第 2 の規格化部 1 6 2 の各値に対し、第 1 の規格化部 1 6 0 の出力する各フレームのサンプリングポイントに対応する値が得られるように線形補間処理を行なって平均計算部 1 6 6 に与える。平均計算部 1 6 6 は、フレームごとに、第 1 の規格化部 1 6 0 の出力と補間処理部 1 6 4 の出力とを正規化し、その結果を時間軸上の非信頼性の分布を示す統合された波形として信頼性の中心候補出力部 1 4 6 に出力する。

信頼性の中心候補出力部 1 4 6 は、凸包アルゴリズムにより、規格化および統合部 1 4 4 の出力する統合された波形の輪郭のディップを検出して、そのフレームを特定する情報を第 3 図の擬似音節中心の抽出部 9 6 に対して信頼性の中心の候補として出力する。

第 3 図に示す擬似音節中心の抽出部 9 6 は、第 6 図に示す信頼性の中心候補出力部 1 4 6 から与えられた信頼性の中心の中で、音響・韻律分析部 9 2 から与えられた擬似音節核の中にあるもののみを擬似音節中心として抽出する。

以上の処理によって、音声データのうちに音声データの特徴を抽出する、または音声データをラベル付けするために適した高信頼性・小変動領域を示す情報が得られたことになる。したがって、この情報によって特定されるフレームについて所望の処理を行なえばよい。本実施例の装置では、擬似音節中心の抽出部 9 6 はこの情報をフォルマントの最適化部 9 8 に与え、フォルマントの最適化部 9 8 はこの情報を用いて、以下のようにして擬似音節中心における A Q 指数を算出する。

なお、本実施例の装置では、擬似音節中心の長さは連続する 5 フレームとする。1 フレームは 3 2 m s e c であり、連続するフレームは互いに 8 m s e c ずつず

れているから、5 フレームの全体では 64 msec の音声期間に相当する。

これらの擬似音節中心における A Q 指数は、第 9 図の A Q 指数計算部 100 中で得られる声門気流の波形により直接計算することができる。しかし、声門気流の推定自体、もともとのフォルマントに相当する声道の共振によって影響されて
5 おり、その信頼性は共振の影響をもとの音声波形の 64 msec のデータから取り除くことができるかに依存している。したがって、そのような計算によって得られた A Q 指数は信頼できないものとなる。

一方、擬似音節中心におけるフォルマントは、スペクトルがよく一致しているという意味で、既によい推定となっているが、本実施例の装置では、さらに以下
10 のようにしてフォルマント周波数を最適化する。

すなわち、第 8 図を参照して、FFT 処理部 180 は音声波形に対してフレームごとに FFT 処理を行なう。対数変換および逆 DCT 部 182 は FFT 処理部 180 の出力に対して対数変換およびコサイン逆変換を行なう。ケプストラム距離計算部 184 は、対数変換および逆 DCT 部 182 の出力するケプストラム係数と距離最小化処理部 186 から与えられるケプストラム係数の推定値との間の距離を計算する。距離最小化処理部 186 は、フォルマントの推定値を表わすケプストラム係数の値を起点として、ケプストラム距離計算部 184 により計算される距離が最小値となるように山登り法によって距離最小化処理部 186 から与えられたケプストラム係数の値をさらに最適化し、最小値が得られるときのフォルマント推定値を出力する。
20

A Q 指数計算部 100 の内部構成は第 9 図に示されており、この第 9 図を参照して、擬似音節中心における音声データはまずハイパスフィルタ 200 を通り、その結果 70 Hz 以下の低周波数の雑音が除去される。さらに適応的ローパスフィルタ 202 によって第 4 フォルマントより高い周波数領域のスペクトル情報が
25 除去される。そして、適応的逆フィルタ 204 によって第 1 ～ 第 4 フォルマントによる影響が除去される。

その結果、適応的逆フィルタ 204 の出力は声門気流の波形の微分のよい推定値となる。これを積分回路 206 で積分することにより声門気流の波形の推定値が得られる。最大ピーク間振幅検出回路 208 によって声門気流の波形のピーク

ツープークの振幅の最大値を検出する。最大の負のピーク振幅検出回路 210 によって声門気流の微分波形のサイクル内での負の最大の振幅を検出する。最大ピーク間振幅検出回路 208 の出力の、最大の負のピーク振幅検出回路 210 の出力に対する比を比計算回路 212 で計算することにより、擬似音節中心における A Q 指数が得られる。

こうして得られた A Q 指数は、各擬似音節中心におけるもとの音声データの特徴（緊張音—氣息音の間の度合い）を信頼性高くあらわしている。これら各擬似音節中心に対して A Q 指数を計算し、さらにこれら得られた A Q 指数を補間することにより、擬似音節中心以外の部分の A Q 指数を推定することもできる。そうすることにより、音声データのうち、一定の A Q 指数を示す部分に、当該 A Q 指数に対応した適切なラベルをパラ言語情報として付けておき、音声合成の際には、所望の A Q 指数を有する音声データを使用すれば、単なるテキストだけでなく、パラ言語情報をも含んだ形での音声合成を行なうことが可能になる。

第 10 図～第 12 図に、本実施例の装置をコンピュータにより実現した際の画面表示例を示す。

第 10 図を参照して、このプログラムによる表示ウィンドウには、音声データ波形 240 と、音声データに対して付された音声ラベル 242 と、基本周波数の波形の時間軸上の分布波形の輪郭 244 と、ソノラントエネルギーの変動の時間軸上の分布波形の輪郭 246 と、 Δ ケプストラムから計算されたスペクトル変化の局所的な分散の時間軸上の分布波形の輪郭 248 と、フォルマントー F F T ケプストラム距離の時間軸上の分布波形の輪郭 250 と、スペクトル変化の局所的な分散の分布波形の輪郭 248 およびフォルマントーケプストラム距離の分布波形の輪郭 250 を統合した波形である非信頼性の時間軸上の分布波形の輪郭 252 と、上述のようにして算出された擬似音節中心での声門の A Q 指数 254 と、各擬似音節中心で推定された声道の面積関数 256 とが示されている。

音声データ波形 240 の表示領域に示された太い縦線 232 と、ソノラントエネルギーの変動の輪郭 246 の表示領域に示された太い縦線とは擬似音節の境界を示す。音声データ波形 240 の表示領域に示された細い縦線 230 と、ソノラントエネルギーの変動の輪郭 246 および基本周波数の波形輪郭 244 の表示領域に

示された細い縦線は擬似音節核の境界を示す。

非信頼性の波形 2 5 2 の表示領域に示された縦線は波形の極小値部分（ディープ）であり、そこを中心として A Q 指数が計算されている部分が最も信頼性の高い部分である。なお A Q 指数が計算された期間および値は横棒で示されており、横棒の縦位置が高いほど緊張音に近く、低いほど気息音に近い。

第 1 1 図には、第 1 0 図の左側の点線のボックス 2 6 2 で示される時点での声門気流波形の推定値 2 7 0 と、その微分波形 2 7 2 と、推定された声門気流波形のスペクトル 2 7 4 とが示されている。第 1 0 図のボックス 2 6 2 に対応する時点では A Q 指数 2 5 4 は高く、すなわちこの時点の発声は緊張音に近い。第 1 1 図に示すとおり、このときの声門気流の波形はのこぎり形に近く、サインウェーブの波形からは遠く異なっている。また、微分波形は鋭く変化している。

第 1 2 図には、第 1 0 図の点線のボックス 2 6 0 で示される時点での声門気流波形の推定値 2 8 0 と、その微分波形 2 8 2 と、推定された声門気流波形のスペクトル 2 8 4 とが示されている。第 1 0 図のボックス 2 6 0 に対応する時点では A Q 指数 2 5 4 は低く、すなわちこの時点の発声は気息音に近い。第 1 2 図に示すとおり、このときの声門気流の波形はきれいなサインカーブに近い。微分波形も緩やかなものとなっている。

上に述べた装置を用い、前述した音声データを実際に処理して擬似音節中心を抽出し、各擬似音節中心に対して A Q 指数を算出する一方、それらの擬似音節中心に対応する音を人間が聞いたときに感ずる感想と、A Q 指数との相関を以下のようにして調査した。

上記した装置を用いて抽出された信頼性の中心は 2 2, 0 0 0 個であり、その各々について対応する声門気流波形および A Q 指数と、もとの音声波形の RMS (Root Mean Square) エネルギー (dB) とを算出した。これら信頼性の中心のうち、同一の音節核中に存在しかつ互いの A Q 指数がほぼ一致しているものをまとめ、さらにそれら信頼性の中心のうち、統合された非信頼性の値が 0. 2 以上のものを棄却することにより、聴覚刺激として使用可能と思われる音節核の数は 1 5, 0 0 0 をわずかに超えたものとなった。

このデータセットに対して算出された統計情報に基づき、知覚上の評価を行な

うために60の刺激からなるサブセットを選択した。具体的には、前述した3つの感情を表わすデータベースの各々について、極めて低い、または極めて高い、または各感情に対するAQ指数の平均値マイナスその分布の標準偏差(σ)近辺、またはAQ指数の平均値プラス標準偏差近辺、の4つのカテゴリのいずれかにAQ指数が属するような信頼性の中心を含む音節核を5つずつ選択した。

このようにして選択された60個の擬似音節核の時間的長さは32 msecから560 msecの範囲であり、その平均は171 msecであった。通常の聴覚的能力を有する11人の被験者が、これら短時間の刺激の各々について聴覚的評価を行なった。被験者は静粛なオフィス環境で、高音質のヘッドフォンを用い、各刺激に必要な回数だけ聞き、各刺激について、それぞれ「気息性」および「強さ」とだけ説明した二つのスケールにしたがい、7段階で採点した。各被験者の採点は各々比例により[0, 1]の範囲に正規化され、正規化した点数に基づいて、60個の刺激の各々についての11人の被験者全ての気息性および強さに関する平均値を算出した。

第13図は、上のようにして調べた気息性と、音響的に測定したAQ指数の値とを比較する散布図である。これら60対の値に対する線形相関係数は0.77であった。この相関は必ずしも高いものではないが、刺激に対するAQの測定値が高くなれば、その刺激に対して感じられる気息性も平均すれば高くなるという明らかな傾向があることを裏付けるものといえる。第13図の散布図上で想定されるベストフィットの直線から最も遠い位置に存在する点のいくつかをより詳細に調べると、誤差の原因として次のようなものが浮かび上がる。すなわち、動的制約が欠如しているために生ずる、5つのフレーム中でのフォルマントの非連続性、5つのフレームに含まれていない音節核の一部において生ずる高い気息性、および5つのフレーム中の母音部分に対して、隣接した鼻音がおよぼす強い影響などである。

さらに、第13図からは、中位から下位のAQ指数を有する刺激に対しては、気息性の感じ方が広いことに気づく。これは、気息性が低い刺激に気息性に関する点数をつけることが難しく、むしろ地声または緊張音的な発音という側面から点数付けしたほうがよりよく特徴を表わせるのではないか、という直感的な理解

を裏付けるものと思われる。

ここでは図としては示していないが、強さの感じ方を、同じ信頼性の中心において測定されたRMSエネルギーと比較するための散布図も作成した。その相関係数は0.83となり、より高度な重み付けを用いて強さの感じ方を測定しているわけではないにもかかわらず、その関係の強さを裏付けるものとなっている。

以上のように本実施例では、音響・韻律的分析と、ケプストラム分析とを組合せて、(i) 録音された自然な発声中の擬似音節の信頼性の中心の位置を決定するための、(ii) 参考文献2で提案されたAQ指数により定量化された音源の属性を測定するための、全くオペレータが介入する必要がない方法および装置を
10 実現した。そして、その方法および装置を用いて行なった音声知覚の実験の結果は、擬似音節核中で知覚された気息性と強い相関を持つ、頑健性をもって測定できる値としてのAQ指数の重要性を確認するものであった。実際、前述したような誤差源が存在しているにもかかわらず、AQ指数と気息性の知覚との間に見出された相関により、音質パラメータとしてのAQ指数をさらに研究する必要があることを確認することができた。

今回開示された実施の形態は単に例示であって、本発明が上記した実施の形態のみに制限されるわけではない。本発明の範囲は、発明の詳細な説明の記載を参酌した上で、特許請求の範囲の各請求項によって示され、そこに記載された文言と均等の意味および範囲内でのすべての変更を含む。

20 [参考文献]

(1) Sundberg, J. (1987). The science of the singing voice, Northern Illinois University Press, Delcalb, Illinois.

(2) Alku, P. & Vilkman, E. (1996). "Amplitude domain quotient for characterization of the glottal volume velocity waveform estimated by
25 inverse filtering", SpeechComm., 18(2), 131-138.

(3) .Hermes, D. (1988). "Measurement of pitch by subharmonic summation", J. Acoust.

Soc. Am. 83(1), 257-264.

(4) Mermelstein, P. (1975). "Automatic segmentation of speech into

syllabic units", J. Acoust. Soc. Am. 58(4), 880-883.

(5) Lea, W.A. (1980). "Prosodic aids to speech recognition", in Lea, W.A. (ed.), Trends in Speech Recognition, Prentice-Hall, New Jersey, 166-205.

5 (6) Broad, D.J. & Clermont, F. (1989). "Formant estimation by linear transformation of the LPC cepstrum", J. Acoust. Soc. Am. 86 (5), 2013-2017.

(7) Mokhtari, P., Iida, A. & Campbell, N. (2001). "Some articulatory correlates of emotion variability in speech : a preliminary study on spoken Japanese vowels", Proc. Int. Conf. on Speech Process., Taejon, Korea, 431-436.

(8) Peterson, G.E., & Shoup, J.E. (1966). "A physiological theory of phonetics", J. Speech Hear. Res. 9, 5-67.

10 (9) Iida, A., Campbell, N., Iga, S., Higuchi, F. & Yasumura, M. (1998). "Acoustic nature and perceptual testing of corpora of emotional speech", Proc. 5th Int. Conf. on Spoken Lang. Process., 1559-1562.

産業上の利用可能性

この方法および装置により、発声単位に対するパラ言語的なラベル付けを、オペレータの介在なく自動的に行なってデータベースを構築することが容易になる。

20 そのように所望のラベル付けがされた発声単位のデータベースを用いて音声の連続合成を行なうことにより、緊張音から地声、さらに気息的な発音までの範囲にわたる幅広い発声スタイルを用いた自然な音声合成を使用したマンマシンインタフェースを実現することが可能となる。

請求の範囲

1. 音声波形のデータに基づいて、前記音声波形の特徴を高い信頼性で示す部分を決定するための装置であって、

前記データから前記音声波形のうちの所定周波数領域のエネルギーの時間軸上の分布を算出し、当該分布および前記音声波形のピッチに基づいて、前記音声波形の各節のうち、前記音声波形の発生源によって安定して発生されている領域を抽出するための抽出手段と、

前記データから前記音声波形のスペクトルの時間軸上の分布を算出し、当該スペクトルの時間軸上の分布に基づいて、前記音声波形のうち、その変化が前記発生源により良好に制御されている領域を推定するための推定手段と、

前記推定手段の出力と、前記発生源によって安定して発生されている領域として前記抽出手段により抽出され、かつ前記発生源によってその変化が良好に制御されていると前記推定手段によって推定された領域を前記音声波形の高精度性部分として決定するための手段とを含む、音声波形の特徴を高い信頼性で示す部分を決定するための装置。

2. 前記抽出手段は、

前記データに基づいて、前記音声波形の各区間が有声区間か否かを判定するための有声判定手段と、

前記音声波形の前記所定周波数領域のエネルギーの時間軸上の分布の波形の極小部で前記音声波形を節に分離するための手段と、

前記音声波形のうち、各節内で、当該節内のエネルギーのピークを含み、かつ前記有声判定手段により有声区間であると判定された区間であって、かつ前記所定周波数領域のエネルギーが所定のしきい値以上である領域を抽出するための手段とを含む、請求項1に記載の装置。

3. 前記推定手段は、

前記音声波形に対する線形予測分析を行ないフォルマント周波数の推定値を出力するための線形予測手段と、

前記データを用いて、前記線形予測手段によるフォルマント周波数の推定値の

非信頼性の時間軸上の分布を算出するための第1の算出手段と、

前記線形予測手段の出力に基づいて、前記音声波形の時間軸上のスペクトル変化の局所的な分散の、時間軸上の分布を算出するための第2の算出手段と、

5 前記第1の算出手段により算出された前記フォルマント周波数の推定値の非信頼性の時間軸上の分布と、前記第2の算出手段により算出された前記音声波形のスペクトル変化の局所的な分散の時間軸上の分布との双方に基づいて、前記音声波形の変化が前記発生源により良好に制御されている領域を推定するための手段を含む、請求項1に記載の装置。

10 4. 前記決定するための手段は、前記推定手段により前記音声波形の変化が前記発生源により良好に制御されていると推定された領域のうち、前記抽出手段により抽出された領域に含まれる領域を前記音声波形の高信頼性部分として決定するための手段を含む、請求項1に記載の装置。

15 5. 音声信号を擬似音節に分離し、さらに各擬似音節の核部分を抽出するための擬似音節核抽出装置であって、

前記音声信号の各区間が有声区間か否かを判定するための有声判定手段と、
前記音声信号の所定周波数領域のエネルギーの時間的な分布の波形の極小部で
前記音声信号を擬似音節に分離するための手段と、

20 前記音声信号のうち、各擬似音節内でのエネルギーのピークを含み、かつ前記有声判定手段により有声区間であると判定された区間である領域を当該擬似音節の核として抽出するための手段を含む、擬似音節核抽出装置。

25 6. 前記抽出するための手段は、前記音声信号のうち、各擬似音節内でのエネルギーのピークを含み、かつ前記有声判定手段により有声区間であると判定された区間であって、かつ前記所定周波数領域のエネルギーが所定のしきい値以上である領域を当該擬似音節の核として抽出するための手段を含む、請求項5に記載の擬似音節核抽出装置。

7. 音声信号の特徴を高い信頼性で示す部分を決定するための装置であって、
前記音声信号に対する線形予測分析を行なうための線形予測手段と、

前記線形予測手段によるフォルマントの推定値と、前記音声信号とに基づいて、
前記フォルマントの推定値の非信頼性の時間軸上の分布を算出するための第1の

算出手段と、

前記線形予測手段による線形予測分析の結果に基づいて、前記音声信号のスペクトル変化の局所的な分散の時間軸上の分布を算出するための第2の算出手段と、

前記第1の算出手段により算出された前記フォルマント周波数の推定値の非信頼性の時間軸上の分布と、前記第2の算出手段により算出された前記音声波形のスペクトル変化の局所的な分散の時間軸上の分布との双方に基づいて、前記音声波形の変化が前記発生源により良好に制御されている領域を推定するための手段とを含む、音声信号の特徴を高い信頼性で示す部分を決定するための装置。

8. コンピュータ上で実行されると、音声波形のデータに基づいて、前記音声波形の特徴を高い信頼性で示す部分を決定するための装置としてコンピュータを動作させるプログラム製品であって、前記装置は、

前記データから前記音声波形のうちの所定周波数領域のエネルギーの時間軸上の分布を算出し、当該分布および前記音声波形のピッチに基づいて、前記音声波形の各節のうち、前記音声波形の発生源によって安定して発生されている領域を抽出するための抽出手段と、

前記データから前記音声波形のスペクトルの時間軸上の分布を算出し、当該スペクトルの時間軸上の分布に基づいて、前記音声波形のうち、その変化が前記発生源により良好に制御されている領域を推定するための推定手段と、

前記推定手段の出力と、前記発生源によって安定して発生されている領域として前記抽出手段により抽出され、かつ前記発生源によってその変化が良好に制御されていると前記推定手段によって推定された領域を前記音声波形の高信頼性部分として決定するための手段とを含む、プログラム製品。

9. 前記抽出手段は、

前記データに基づいて、前記音声波形の各区間が有声区間か否かを判定するための有声判定手段と、

前記音声波形の前記所定周波数領域のエネルギーの時間軸上の分布の波形の極小部で前記音声波形を節に分離するための手段と、

前記音声波形のうち、各節内で、当該節内のエネルギーのピークを含み、かつ前記有声判定手段により有声区間であると判定された区間であって、かつ前記所

定周波数領域のエネルギーが所定のしきい値以上である領域を抽出するための手段とを含む、請求項 8 に記載のプログラム製品。

10. 前記推定手段は、

前記音声波形に対する線形予測分析を行ないフォルマント周波数の推定値を出力するための線形予測手段と、

前記データを用いて、前記線形予測手段によるフォルマント周波数の推定値の非信頼性の時間軸上の分布を算出するための第 1 の算出手段と、

前記線形予測手段の出力に基づいて、前記音声波形の時間軸上のスペクトル変化の局所的な分散の、時間軸上の分布を算出するための第 2 の算出手段と、

前記第 1 の算出手段により算出された前記フォルマント周波数の推定値の非信頼性の時間軸上の分布と、前記第 2 の算出手段により算出された前記音声波形のスペクトル変化の局所的な分散の時間軸上の分布との双方に基づいて、前記音声波形の変化が前記発生源により良好に制御されている領域を推定するための手段とを含む、請求項 8 に記載のプログラム製品。

11. 前記決定するための手段は、前記推定手段により前記音声波形の変化が前記発生源により良好に制御されていると推定された領域のうち、前記抽出手段により抽出された領域に含まれる領域を前記音声波形の高信頼性部分として決定するための手段を含む、請求項 8 に記載のプログラム製品。

12. コンピュータ上で実行されると、音声信号を擬似音節に分離し、さらに各擬似音節の核部分を抽出するための擬似音節核抽出装置としてコンピュータを動作させるプログラム製品であって、前記擬似音節核抽出装置は、

前記音声信号の各区間が有声区間か否かを判定するための有声判定手段と、

前記音声信号の所定周波数領域のエネルギーの時間的な分布の波形の極小部で前記音声信号を擬似音節に分離するための手段と、

前記音声信号のうち、各擬似音節内でのエネルギーのピークを含み、かつ前記有声判定手段により有声区間であると判定された区間であって、かつ前記所定周波数領域のエネルギーが所定のしきい値以上である領域を当該擬似音節の核として抽出するための手段とを含む、プログラム製品。

13. 音声信号の特徴を高い信頼性で示す部分を決定するための装置として

コンピュータを動作させるプログラム製品であって、前記装置は、

前記音声信号に対する線形予測分析を行なうための線形予測手段と、

前記線形予測手段によるフォルマントの推定値と、前記音声信号とに基づいて、
前記フォルマントの推定値の非信頼性の時間軸上の分布を算出するための第1の
5 算出手段と、

前記線形予測手段による線形予測分析の結果に基づいて、前記音声信号のスペ
クトル変化の局所的な分散の時間軸上の分布を算出するための第2の算出手段と、

前記第1の算出手段により算出された前記フォルマント周波数の推定値の非信
頼性の時間軸上の分布と、前記第2の算出手段により算出された前記音声波形の
10 スペクトル変化の局所的な分散の時間軸上の分布との双方に基づいて、前記音声
波形の変化が前記発生源により良好に制御されている領域を推定するための手段
とを含む、音声信号の特徴を高い信頼性で示す部分を決定するためのプログラム
製品。

14. 音声波形のデータに基づいて、前記音声波形の特徴を高い信頼性で示
15 す部分を決定する方法であって、

前記データから前記音声波形のうちの所定周波数領域のエネルギーの時間軸上
の分布を算出し、当該分布および前記音声波形のピッチに基づいて、前記音声波
形の各節のうち、前記音声波形の発生源によって安定して発生されている領域を
抽出するステップと、

20 前記データから前記音声波形のスペクトルの時間軸上の分布を算出し、当該ス
ペクトルの時間軸上の分布に基づいて、前記音声波形のうち、その変化が前記発
生源により良好に制御されている領域を推定するステップと、

前記推定するステップの出力と、前記発生源によって安定して発生されている
領域として前記抽出するステップにおいて抽出され、かつ前記発生源によってそ
25 の変化が良好に制御されていると前記推定するステップにおいて推定された領域
を前記音声波形の高精度部分として決定するステップとを含む、方法。

15. 前記抽出するステップは、

前記データに基づいて、前記音声波形の各区間が有声区間か否かを判定するス
テップと、

前記音声波形の前記所定周波数領域のエネルギーの時間軸上の分布の波形の極小部を検出し、当該極小部で前記音声波形を節に分離するステップと、

前記音声波形のうち、各節内で、当該節内のエネルギーのピークを含み、かつ前記有声判定手段により有声区間であると判定された区間であって、かつ前記所定周波数領域のエネルギーが所定のしきい値以上である領域を抽出するステップ

5 とを含む、請求項 1 4 に記載の方法。

1 6. 前記推定するステップは、

前記音声波形に対する線形予測分析を行ないフォルマント周波数の推定値を出力するステップと、

10 前記データを用いて、前記推定値を出力するステップにより推定されたフォルマント周波数の非信頼性の時間軸上の分布を算出するステップと、

算出された前記フォルマント周波数の非信頼性の時間軸上の分布に基づいて、前記音声波形の時間軸上のスペクトル変化の局所的な分散の、時間軸上の分布を算出するステップと、

15 前記算出された前記フォルマント周波数の非信頼性の時間軸上の分布と、算出された前記音声波形のスペクトル変化の局所的な分散の時間軸上の分布との双方に基づいて、前記音声波形の変化が前記発生源により良好に制御されている領域を推定するステップとを含む、請求項 1 4 に記載の方法。

1 7. 前記決定するステップは、前記推定するステップにおいて前記音声波形の変化が前記発生源により良好に制御されていると推定された領域のうち、前記抽出するステップにより抽出された領域に含まれる領域を前記音声波形の高信頼性部分として決定するステップを含む、請求項 1 4 に記載の方法。

1 8. 音声信号を擬似音節に分離し、さらに各擬似音節の核部分を抽出する方法であって、

25 前記音声信号の各区間が有声区間か否かを判定するステップと、

前記音声信号の所定周波数領域のエネルギーの時間的な分布の波形の極小部で前記音声信号を擬似音節に分離するステップと、

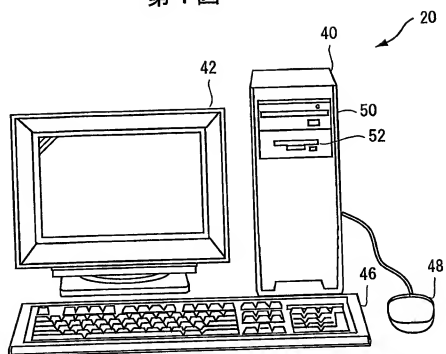
前記音声信号のうち、各擬似音節内でのエネルギーのピークを含み、かつ前記判定するステップにおいて有声区間であると判定された区間である領域を当該擬

似音節の核として抽出するステップとを含む、方法。

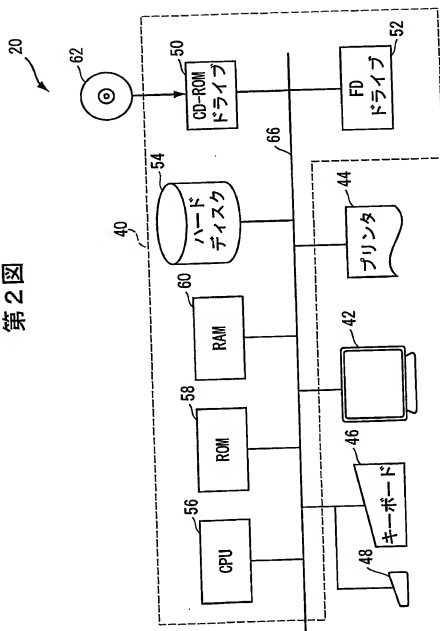
19. 前記抽出するステップは、前記音声信号のうち、各擬似音節内でのエネルギーのピークを含み、かつ前記判定するステップにおいて有声区間であると判定された区間であって、かつ前記所定周波数領域のエネルギーが所定のしきい値以上である領域を当該擬似音節の核として抽出するステップを含む、請求項18に記載の方法。

5

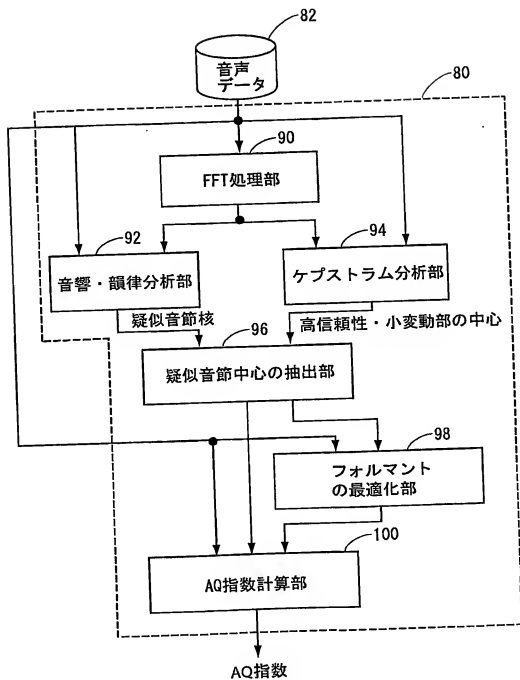
第 1 図



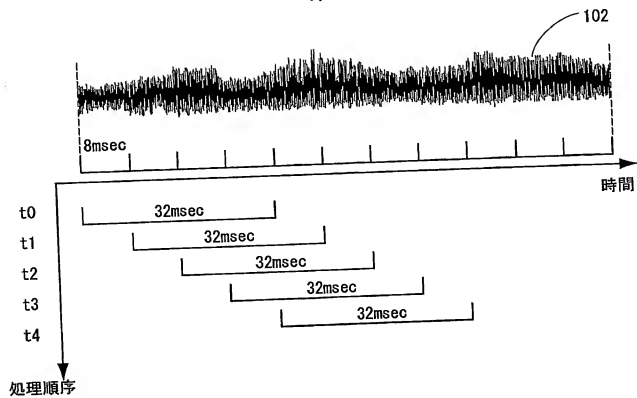
第2図



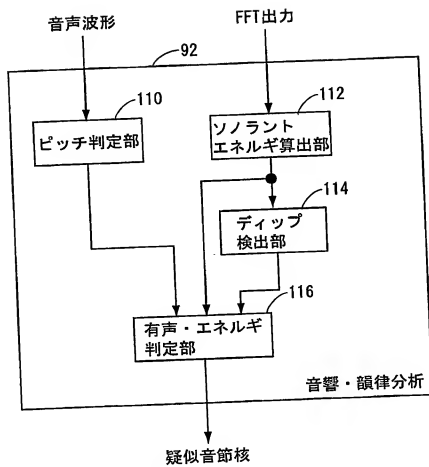
第3図



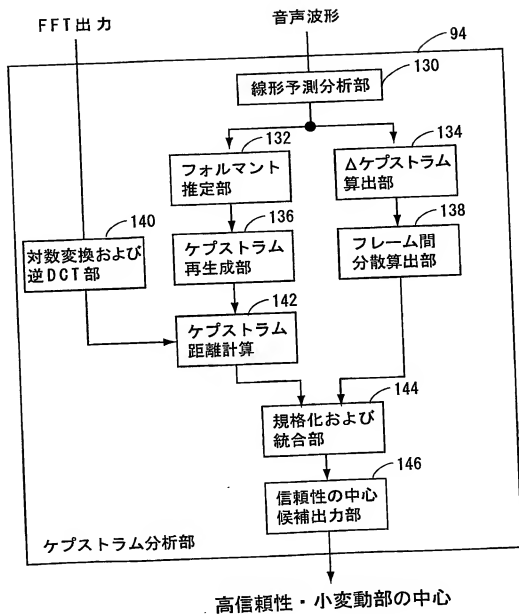
第 4 図



第5図

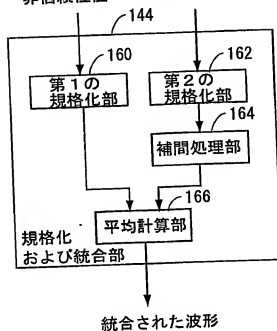


第6図

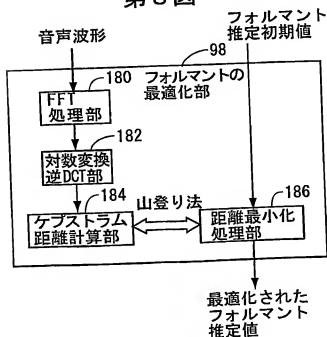


第7図

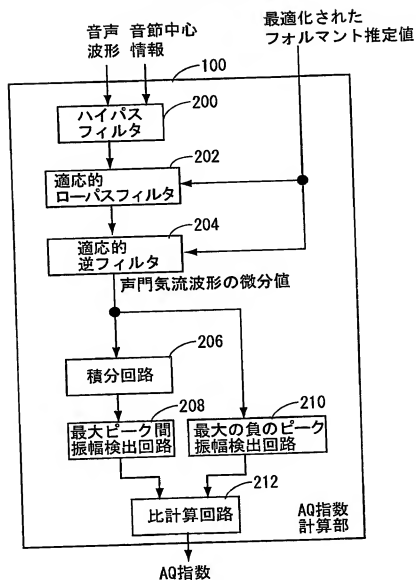
フレームごとの 局所的な
非信頼性値 フレーム間分散



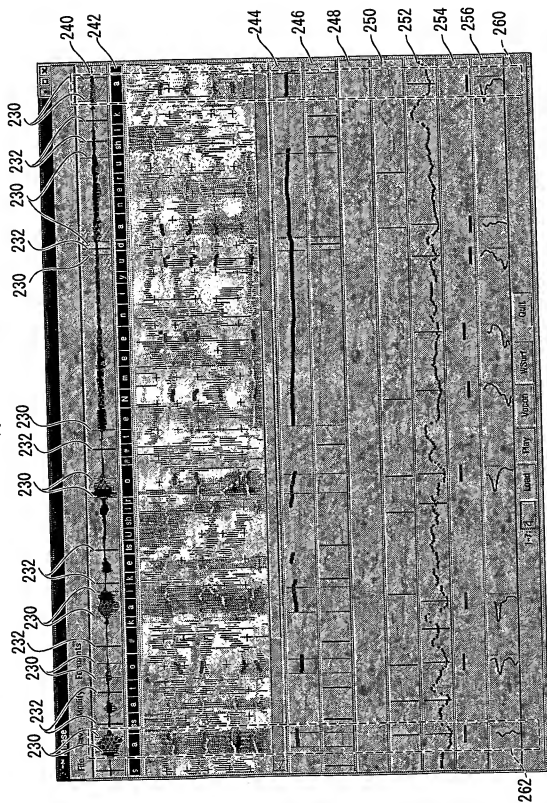
第8図



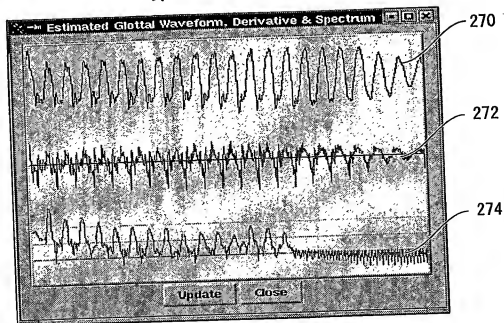
第9図



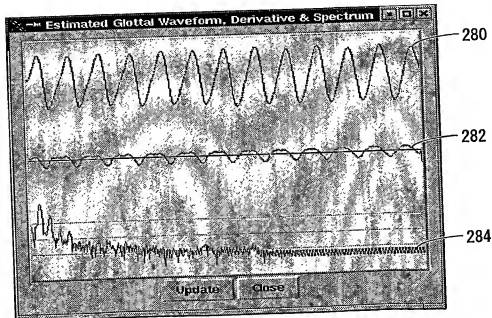
第10図



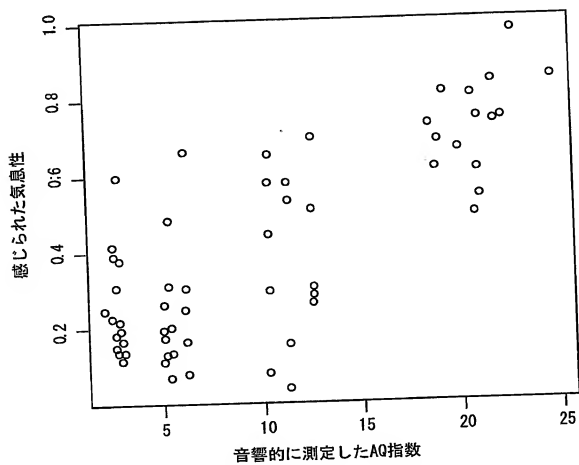
第 1 1 図



第 1 2 図



第 13 図



特許協力条約に基づく国際出願願書

原本（出願用） - 印刷日時 2003年02月19日（19. 02. 2003）水曜日 16時28分58秒

VIII-5-1	不利にならない開示又は新規性喪失の例外に関する申立て 不利にならない開示又は新規性喪失の例外に関する申立て（規則4.17(v)及び51の2.1(a)(v)）	本国際出願に関し、 科学技術振興事業団は、本国際出願の請求項に記載された対象が以下のよう に開示されたことを申し立てる。
VIII-5-1	開示の種類	刊行物
(i)		
VIII-5-1	開示の日付：	2002年03月18日（18. 03. 2002）
(ii)		
VIII-5-1	開示の名称：	日本音響学会2002年春季研究発表会講演論文集- I
(iii)		
VIII-5-1	開示の場所：	神奈川大学
(iv)		
VIII-5-1	本申立ては、次の指定国のため	国内特許又は広域特許のための CA の指定
(v)	になされたものである。：	

特許協力条約に基づく国際出願願書

原本（出願用） - 印刷日時 2003年02月19日（19. 02. 2003）水曜日 16時28分58秒

VIII-5-2	不利にならない開示又は新規性喪失の例外に関する申立て 不利にならない開示又は新規性喪失の例外に関する申立て（規則4.17(v)及び51の2.1(a)(v)）	本国際出願に関し、 株式会社国際電気通信基礎技術研究所は、本国際出願の請求項に記載された対象が以下のよう に開示されたことを申し立てる。
VIII-5-2	開示の種類	刊行物
(i) VIII-5-2	開示の日付:	2002年03月18日（18. 03. 2002）
(ii) VIII-5-2	開示の名称:	日本音響学会2002年春季研究発表会講演論文集- I
(iii) VIII-5-2	開示の場所:	神奈川大学
(iv) VIII-5-2	本申立ては、次の指定国のため	国内特許又は広域特許のための CA の指定
(v)	になされたものである。:	

特許協力条約に基づく国際出願願書

原本（出願用） - 印刷日時 2003年02月19日（19.02.2003）水曜日 16時28分58秒

VIII-5-3	不利にならない開示又は新規性喪失の例外に関する申立て 不利にならない開示又は新規性喪失の例外に関する申立て（規則4.17(v)及び51の2.1(a)(v)）	本国際出願に関し、 キャンベル ニックは、 本国際出願の請求項に記載された対象が以下のよう に開示されたことを申し立てる。
VIII-5-3 (i) VIII-5-3 (ii) VIII-5-3 (iii)	開示の種類 開示の日付： 開示の名称： 開示の場所：	刊行物 2003年03月18日（18.03.2003） 日本音響学会2002年春季研究発表会講演論文集-I - 神奈川大学
VIII-5-3 (v)	本申立ては、次の指定国のため になされたものである。：	国内特許又は広域特許のための US の指定

特許協力条約に基づく国際出願願書

原本（出願用） - 印刷日時 2003年02月19日（19.02.2003）水曜日 16時28分58秒

VIII-5-4	不利にならない開示又は新規性喪失の例外に関する申立て 不利にならない開示又は新規性喪失の例外に関する申立て（規則4.17(v)及び51の2.1(a)(v)）	本国際出願に関し、 モクタリ パーハムは、 本国際出願の請求項に記載された対象が以下のよう に開示されたことを申し立てる。
VIII-5-4 (i)	開示の種類	刊行物
VIII-5-4 (ii)	開示の日付:	2002年03月18日（18.03.2002）
VIII-5-4 (iii)	開示の名称:	日本音響学会2002年春季研究発表会講演論文集-I
VIII-5-4 (iv)	開示の場所:	神奈川大学
VIII-5-4 (v)	本申立ては、次の指定国のため になされたものである。:	国内特許又は広域特許のための US の指定

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP03/01954

A. CLASSIFICATION OF SUBJECT MATTER
Int.Cl.⁷ G10L11/00, G10L13/06

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
Int.Cl.⁷ G10L11/00, G10L13/06

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Jitsuyo Shinan Koho	1922-1996	Toroku Jitsuyo Shinan Koho	1994-2003
Kokai Jitsuyo Shinan Koho	1971-2003	Jitsuyo Shinan Toroku Koho	1996-2003

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
JICST (JOIS)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP 1-244499 A (Toshiba Corp.), 28 September, 1989 (28.09.89), (Family: none)	1-19
A	JP 10-260697 A (Matsushita Electric Industrial Co., Ltd.), 29 September, 1998 (29.09.98), (Family: none)	1-19
A	JP 2001-282277 A (Canon Inc.), 12 October, 2001 (12.10.01), & US 2002/51955 A1	1-19
A	JP 2001-306087 A (Ricoh Co., Ltd.), 02 November, 2001 (02.11.01), (Family: none)	1-19

☒ Further documents are listed in the continuation of Box C.

☐ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier document but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed

"I" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search
01 April, 2003 (01.04.03)

Date of mailing of the international search report
15 April, 2003 (15.04.03)

Name and mailing address of the ISA/
Japanese Patent Office

Authorized officer

Facsimile No.

Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP03/01954

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	Mitsuo FUJIMOTO, Toru KITAMURA, Masayuki IIDA, Ryuji SUZUKI, "Onsei Suhen Sakusei Shien System", The Institute of Electronics, Information and Communication Engineers Zenkoku Taikai Koen Ronbunshu, 1989, Vol.1989, No.Spring, Pt.1, page 14	1-19

A. 発明の属する分野の分類 (国際特許分類 (IPC))

Int. Cl⁷ G10L11/00, G10L13/06

B. 調査を行った分野

調査を行った最小限資料 (国際特許分類 (IPC))

Int. Cl⁷ G10L11/00, G10L13/06

最小限資料以外の資料で調査を行った分野に含まれるもの

日本国実用新案公報 1922-1996年
 日本国公開実用新案公報 1971-2003年
 日本国登録実用新案公報 1994-2003年
 日本国実用新案登録公報 1996-2003年

国際調査で使用した電子データベース (データベースの名称、調査に使用した用語)

JICST (JOIS)

C. 関連すると認められる文献

引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求の範囲の番号
A	J P 1-244499 A (株式会社東芝), 1989.09.28 (ファミリーなし)	1-19
A	J P 10-260697 A (松下電器産業株式会社), 1998.09.29 (ファミリーなし)	1-19
A	J P 2001-282277 A (キャノン株式会社), 2001.10.12 & US 2002/51955 A1	1-19
A	J P 2001-306087 A (株式会社リコー),	1-19

☒ C欄の続きにも文献が列挙されている。☐ パテントファミリーに関する別紙を参照。

* 引用文献のカテゴリー

「A」特に関連のある文献ではなく、一般的技术水準を示すもの

「E」国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの

「L」優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献 (理由を付す)

「O」口頭による開示、使用、展示等に言及する文献

「P」国際出願日前で、かつ優先権の主張の基礎となる出願

の日の後に公表された文献

「T」国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの

「X」特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの

「Y」特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの

「&」同一パテントファミリー文献

国際調査を完了した日

01.04.03

国際調査報告の発送日

15.04.03

国際調査機関の名称及びあて先

日本国特許庁 (ISA/J P)

郵便番号100-8915

東京都千代田区霞が関三丁目4番3号

特許庁審査官 (権限のある職員)

渡邊 聡



5C 8622

電話番号 03-3581-1101 内線 3540

C (続き) . 関連すると認められる文献		関連する
引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	請求の範囲の番号
A	2001.11.02 (ファミリーなし) 藤本光男, 北村徹, 飯田正幸, 鈴木龍司, 音声素片作成支援システム, 電子情報通信学会全国大会講演論文集, 1989, Vol. 1 989, No. Spring, Pt. 1, PAGE. 14	1-19